dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS HVH, phs000993.v5.p2
"*NHLBI TOPMed: Heart and Vascular Health Study (HVH)*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

October	19, 2016	Version 1 Data set release date
October	26, 2017	Version 2 Data set release date
January	23, 2018	Update VCFs – chr3
January	9, 2019	Version 3 Data set release date
October	17, 2019	Version 4 Data set release date
June	2, 2021	Version 5 Data set release date

2021-06-02
Version 5 Data set release for NHLBI TOPMed WGS HVH now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

**There are no overlapping subjects between the 2 consent groups listed below.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, MDS) (HMB-IRB-MDS)

Data Type	subjects	samples
Phenotype	697	699
Seq_DNA_SNP_CNV (VCFs)	685	687
WGS*	685	687

Consent group 2 (c2): Disease-Specific (Cardiovascular Disease, IRB, MDS) (DS-CVD-IRB-MDS)

Data Type	subjects	samples
Phenotype	12	12
Seq_DNA_SNP_CNV (VCFs)	12	12
WGS*	12	12

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see: https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS HVH version 3 phs000993.v4.p2 has been updated to version 5. The dbGaP accession for the current data is **phs000993.v5.p2**. The participant number (p#) has not changed in version 5. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

1. New genotype data of version 5 are accessioned under phg001565.v1. Please see phg001565.v1.TOPMed_WGS_HVH_v5_frz9.sample-info.MULTI.tar.gz folder for genotyped samples, consent status and sample mapping file.
2. Please see phg001565.v1.TOPMed_WGS_HVH_v5_frz9.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-IRB-MDS.tar.gz and phg001565.v1.TOPMed_WGS_HVH_v5_frz9.genotype-calls-vcf.WGS_markerset_grc38.c2.DS-CVD-IRB-MDS.tar.gz for molecular data in form of vcf files.
3. The submitter provided QC files for this data are in the folder marked as phg001565.v1.TOPMed_WGS_HVH_v5_frz9.genotype-qc.MULTI.tar.gz.

phg001277.v1	Freeze 8
phg001565.v1	Freeze 9

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000993/phs000993.v5.p2

2019-10-17

Version 4 Data set release for NHLBI TOPMed WGS HVH now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

**There are no overlapping subjects between the 2 consent groups listed below.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, MDS) (HMB-IRB-MDS)

Data Type	subjects	samples
Phenotype	697	699
Seq_DNA_SNP_CNV (VCFs)	685	687
WGS*	685	687

Consent group 2 (c2): Disease-Specific (Cardiovascular Disease, IRB, MDS) (DS-CVD-IRB-MDS)

Data Type	subjects	samples
Phenotype	12	12
Seq_DNA_SNP_CNV (VCFs)	12	12
WGS*	12	12

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see: https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS HVH version 3 phs000993.v3.p2 has been updated to version 4. The dbGaP accession for the current data is **phs000993.v4.p2**. The participant number (p#) has not changed in version 4. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

1. New genotype data of version 4 are accessioned under phg001277.v1. Please see phg001277.v1.TOPMed_WGS_HVH_v4.sample-info.MULTI.tar.gz folder for genotyped samples, consent status, and sample mapping file.
2. Please see phg001277.v1.TOPMed_WGS_HVH_v4.genotype-calls-vcf.WGS_markerset_grc38.c1.HMB-IRB-MDS.tar.gz for molecular data in form of vcf files.
3. The submitter provided QC files for this data are in the folder marked as phg001277.v1.TOPMed_WGS_HVH_v4.genotype-qc.MULTI.tar.gz.
4. Only the VCFs from the 2 most current freezes will be available for download.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000993/phs000993.v4.p2

2019-01-09
Version 3 Data set release for NHLBI TOPMed WGS HVH now available

dbGaP Study Release Notes

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

**There are no overlapping subjects between the 2 consent groups listed below.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, MDS) (HMB-IRB-MDS)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	WGS
subjects	697	602	602
samples	699	604	604

Consent group 2 (c2): Disease-Specific (Cardiovascular Disease, IRB, MDS) (DS-CVD-IRB-MDS)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	WGS
subjects	12	10	10
samples	12	10	10

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS HVH version 2 phs000993.v2.p2 has been updated to version 3. The dbGaP accession for the current data is **phs000993.v3.p2**. The participant number (p#) has not changed in version 3. No new subjects have been added to this study.

2. **Updated Datasets (n=1)**

pht	version	Dataset Name
5014	3	TOPMed_WGS_HVH_Sample

3. Please note we are discontinuing the submission and distribution of the SAMPLE_USE variable. The sample use counts will be populated by SRA (sequences) and dbGaP (all other submitted molecular data).

Molecular Data Updates

Genotype data are accessioned under phg001152.v1:

1. Please see “sample-info” component for genotyped samples, sample consent status, and mapping of sample-vcf file.
2. Genotypes from whole genome sequencing of 614 samples are available in originally submitted variant-call-format (VCFv4.2). They are marked as “genotype-calls-vcf” and packed into two folders by subject consent status. The reference genome build is GRCh38.
3. QC results, including both submitter provided quality data and consistency-checking results from dbGaP, are in the folder “phg001152.v1.TOPMed_WGS_HVH_v3.genotype-qc.WGS_markerset_grc38.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000993/phs000993.v3.p2

2018-01-23
Update VCFs – chr3

TOPMed WGS_HVH genotypes (vcfs) are updated to phg000960.v2 by adding back chromosome 3 vcf files. There is no change in sample info or consent status.

2017-10-26
Version 2 Data set release for NHLBI TOPMed WGS HVH now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

**There are no overlapping subjects between the 2 consent groups listed below.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, MDS) (HMB-IRB-MDS)

	Phenotype	Seq_DNA_SNP_CNV	Seq_DNA_WholeGenome
subjects	697	71	71
samples	699	73	73

Consent group 2 (c2): Disease-Specific (Cardiovascular Disease, IRB, MDS) (DS-CVD-IRB-MDS)

	Phenotype	Seq_DNA_SNP_CNV	Seq_DNA_WholeGenome
subjects	12	4	4
samples	12	4	4

Molecular data descriptions:
(https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi)

a. Seq_DNA_SNP_CNV: SNP and CNV genotypes derived from sequence data (VCFs)
b. Seq_DNA_WholeGenome: Whole genome sequencing

Study and Phenotype Updates

1. **New Study Accession**

NHLBI TOPMed WGS HVH version 1 phs000993.v1.p1 has been updated to version 2. The dbGaP accession for the current data is **phs000993.v2.p2**. The participant number (p#) has changed in version 2; subjects have been retired. There are new subjects that have been added.

2. **Updated Datasets (n=3; all existing variables have been updated)**

pht	version	Dataset Name
5013	2	TOPMed_WGS_HVH_Subject
5014	2	TOPMed_WGS_HVH_Sample
5015	2	TOPMed_WGS_HVH_Sample_Attributes

3. **New Variables (n=5)**

pht	pht version	Dataset Name	phv	Variable Name
5015	2	TOPMed_WGS_HVH_Sample_Attributes	314692	SEQUENCING_CENTER
5015	2	TOPMed_WGS_HVH_Sample_Attributes	314693	Funding_Source
5015	2	TOPMed_WGS_HVH_Sample_Attributes	314694	TOPMed_Phase
5015	2	TOPMed_WGS_HVH_Sample_Attributes	314695	TOPMed_Project
5015	2	TOPMed_WGS_HVH_Sample_Attributes	314696	Study_Name

Molecular Data Updates

1. Genotype data are accessioned under phg000960.v1. Please see “sample-info” component for genotyped samples, sample consent status and mapping of sample-vcf file.
2. Genotypes from whole genome sequencing are available in originally submitted variant-call-format (VCFv4.2). The “genotype-calls-vcf” are packed into two folders split by subject consent status.
3. QC results, including both submitter provided quality data and consistency-checking results from dbGaP, are in the folder “phg000960.v1.TOPMed_WGS_VAFAR_v2.genotype-qc.WGS_markerset_grc37.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000993/phs000993.v2.p2

dbGaP Study Release Notes

Version 1 Data set release for NHLBI TOPMed WGS HVH now available

This release includes TOPMed Phase I phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

**There are no overlapping subjects between the 2 consent groups listed below.

Consent group 1 (c1): Health/Medical/Biomedical (IRB, MDS) (HMB-IRB-MDS)

	phenotype	SRA/VCFs
subjects	74	68
samples	75	69

Consent group 2 (c2): Disease-Specific (Cardiovascular Disease, IRB, MDS) (DS-CVD-IRB-MDS)

	phenotype	SRA/VCFs
subjects	4	4
samples	4	4

Molecular Data Updates

1. Genotype data are accessioned under phg000800.v1. Please see “sample-info” component for genotyped samples, sample consent status and mapping of sample-vcf file.
2. Genotypes from whole genome sequencing are available in originally submitted variant-call-format (VCFv4.2). They are packed into two folders, marked as “genotype-calls-vcf”, on sample consent status.
3. QC results, including both submitter provided quality data and consistency-checking results by dbGaP, are in the folder “phg000800.v1.TOPMed_WGS_HVH.genotype-qc.WGS_markerset_grc37.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000993/phs000993.v1.p1